

STATISTICAL APPROACH FOR DISCRIMINATION OF BACKGROUND AND IMPACTED AREAS FOR MIDNITE MINE RI/FS

Prepared for
U.S. Environmental Protection Agency
Region 10

October 5, 2001



URS Corporation
8181 East Tufts Avenue
Denver, CO 80237

Project No. 53F4001800.06 04100

Table of Contents

<u>Section</u>	<u>Page</u>
1.0 INTRODUCTION	1
2.0 EVALUATION OF BACKGROUND DATA	6
2.1 Grouping of Data by Medium	7
2.1.1 Groundwater	7
2.1.2 Surface Water.....	8
2.1.3 Sediment.....	8
2.1.4 Surface Material.....	9
2.2 Handling Parameter Values Below Detection Limit.....	9
2.2.1 Helsel's Robust Method Applied to Samples Below the Detection Limit.....	6
2.2.2 2 Sigma Values from Radiological Laboratory Analyses	10
2.2.3 Negative ES Values from Radiological Laboratory Analyses.....	10
2.3 Univariate Statistics and Distribution Test	11
2.3.1 Basic Statistics.....	11
2.3.2 Shapiro-Wilk Distribution Test.....	16
2.4 Selection of Indicator Parameters.....	16
2.4.1 Correlation Coefficients and Matrix.....	16
2.4.2 Professional Judgement Assessment	17
3.0 COMPARISON OF PIA SAMPLES TO BACKGROUND LEVELS	18
3.1 Exceedance Wheel Plots	19
3.2 Mann-Whitney "U" for Surface Material Indicator Parameters	21
3.3 Multivariate Graphical Assessment.....	22
3.4 Final Discrimination and Plotting Locations.....	22
4.0 STATISTICAL CHARACTERIZATION OF MINE-AFFECTED SUB-AREA POPULATIONS.....	23
4.1 Calculation of Univariate Statistics.....	23
4.2 Graphical Comparisons of Populations.....	23

Table of Contents

4.3 Analysis of Parameter Values in Affected Sub-areas.....	24
5.0 REFERENCES	25

LIST OF TABLES

Table 1	Example .01 File from SAS Output
Table 2	Chart with Factor λ for both a 99% UTL and a 95% UPL

LIST OF FIGURES

Figure 1	Logic Flow Diagram 1
Figure 2	Logic Flow Diagram 2
Figure 3	Logic Flow Diagram 3
Figure 4	Example Exceedance Wheel Plot Showing 95% UPL for Water Samples with $k = 2$ UTL Exceedances but no UPL Exceedance
Figure 5	Example Exceedance Wheel Plot Showing 95% UPL for Water Samples which have $k = 2$ UTL Exceedances and 1 UPL Exceedance

1.0 INTRODUCTION

The purpose of this document is to describe the technical approach, scope, rationale, and methods for the statistical analyses needed to support the Remedial Investigation (RI) and the Feasibility Study (FS) for the Midnite Mine Superfund Site. This statistical approach technical memo (SATM) represents a refinement and expansion of the statistical approach previously described in the Quality Assurance Project Plan (QAPP) documents (URS Greiner 1999 and URS 2000). This SATM describes the activities involved in statistically evaluating whether or not media at specific locations and areas near the site have been impacted by mining activities. These results will be used in defining the nature and extent of mine-affected materials and water for the RI.

A large number of chemical and physical measurements in media such as groundwater (GW samples), surface water (SW samples), sediment (SED samples), and surface material (SM samples) have been collected at the Midnite Mine site as described in the Phase 1A/2A-1B QAPPs. This SATM augments the description of the statistical methods that will be employed to evaluate the data. The primary objective of this statistical work is to identify areas impacted by mining as compared to background levels. A secondary objective of this work is to assist in characterizing constituent levels in the affected media in a manner that will facilitate further statistical analysis of the data for the risk assessments.

While the work described herein is focused on discriminating potentially impacted area (PIA) samples that have been affected by mining from those that have not, the background limit (BL) values will also be considered in evaluating levels of contamination within the mined area (MA) as part of the RI report.

Characterization of the concentrations of naturally-occurring inorganic constituents is needed to evaluate whether groundwater, surface water, stream-deposited sediment or mechanically or wind-blown surface material in the PIA have been affected or contaminated by materials and activities in the MA. For all media, background (BG) is defined as the range of chemical/radiological concentrations that are naturally occurring in the site vicinity, in areas unaffected by the previous mining operations. Ideally, such an evaluation would be based on the comparison of existing environmental conditions at, and in the vicinity of, the mine to pre-mining conditions in the same area. Therefore, in order to characterize background conditions for this RI/FS, sampling and measuring natural constituent concentrations occurred at the following locations:

- Near the Mined Area (the area of actual disturbance caused by mining activities) and are near, but not within, the PIA (undisturbed but potentially impacted areas surrounding the MA)
- Similar hydrogeologic characteristics to the Midnite Mine
- Not affected by mining
- Accessible for drilling and sampling activities

Based on these criteria, locations were selected, sampled, and analyzed for radiological parameters and chemical constituent concentrations that represent natural background levels in a geologic setting analogous to that in the MA and PIA. The methodology and rationale for the background sampling are provided in the Phase 1A/2A-1B QAPP documents (URS Greiner 1999 and URS 2000).

In overview, the statistical approach for this study consists of these three major components:

1. Evaluation of background data
2. Comparison of PIA samples to background levels
3. Statistical characterization of parameter values in mine-affected sub-areas

The sequence of work activities is illustrated on logic flow diagrams (Figures 1 through 3). Logic Flow Diagram 1 (Figure 1) shows the steps for the background statistics, culminating with the calculation of BLs for all measured parameters. Logic Flow Diagram 2 (Figure 2) shows the process for comparing BLs with PIA sample data and discriminating samples that have been affected by the mining. The sample discrimination process ends at the bottom of Figure 2. Logic Flow Diagram 3 (Figure 3) illustrates the statistical analyses that will be performed to characterize the nature of contamination within the affected sub-areas. Details of each major component of this study, including the purpose, scope, methodology and rationale, are presented in the following sections.

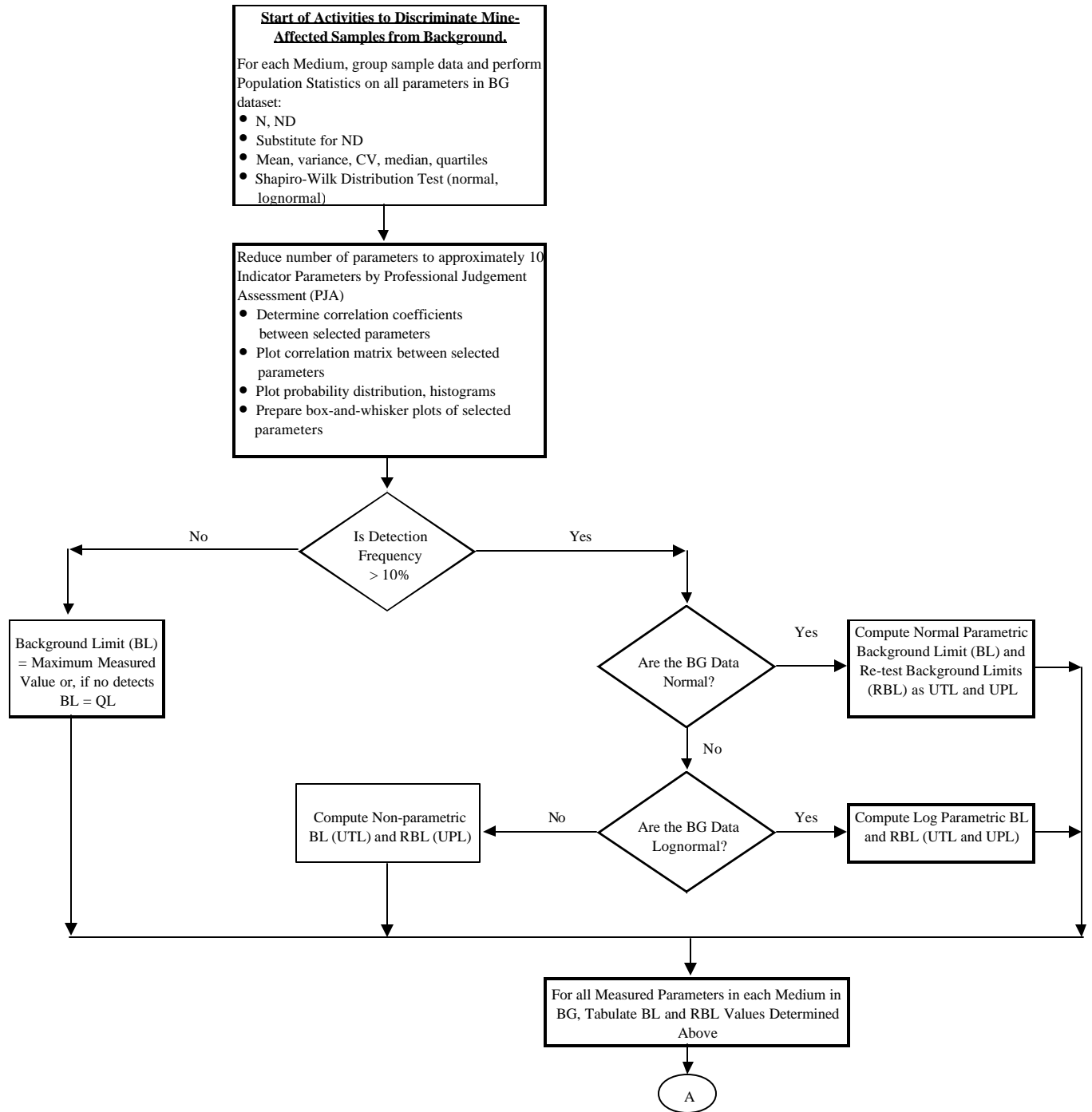
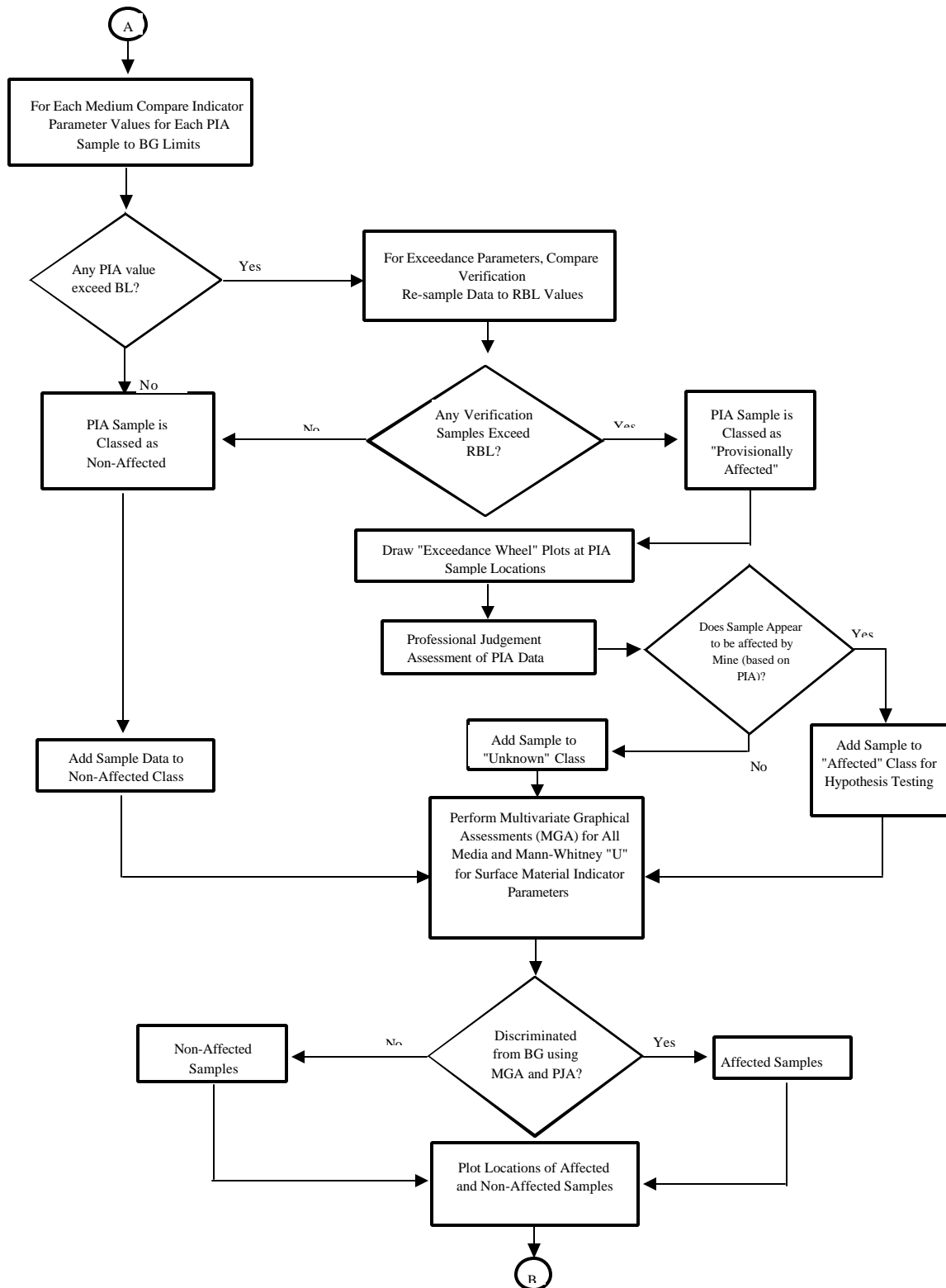


Figure 1: Logic Flow Diagram 1



[End of Activities to Discriminate Mine-Affected Samples from Background]

Figure 2: Logic Flow Diagram 2

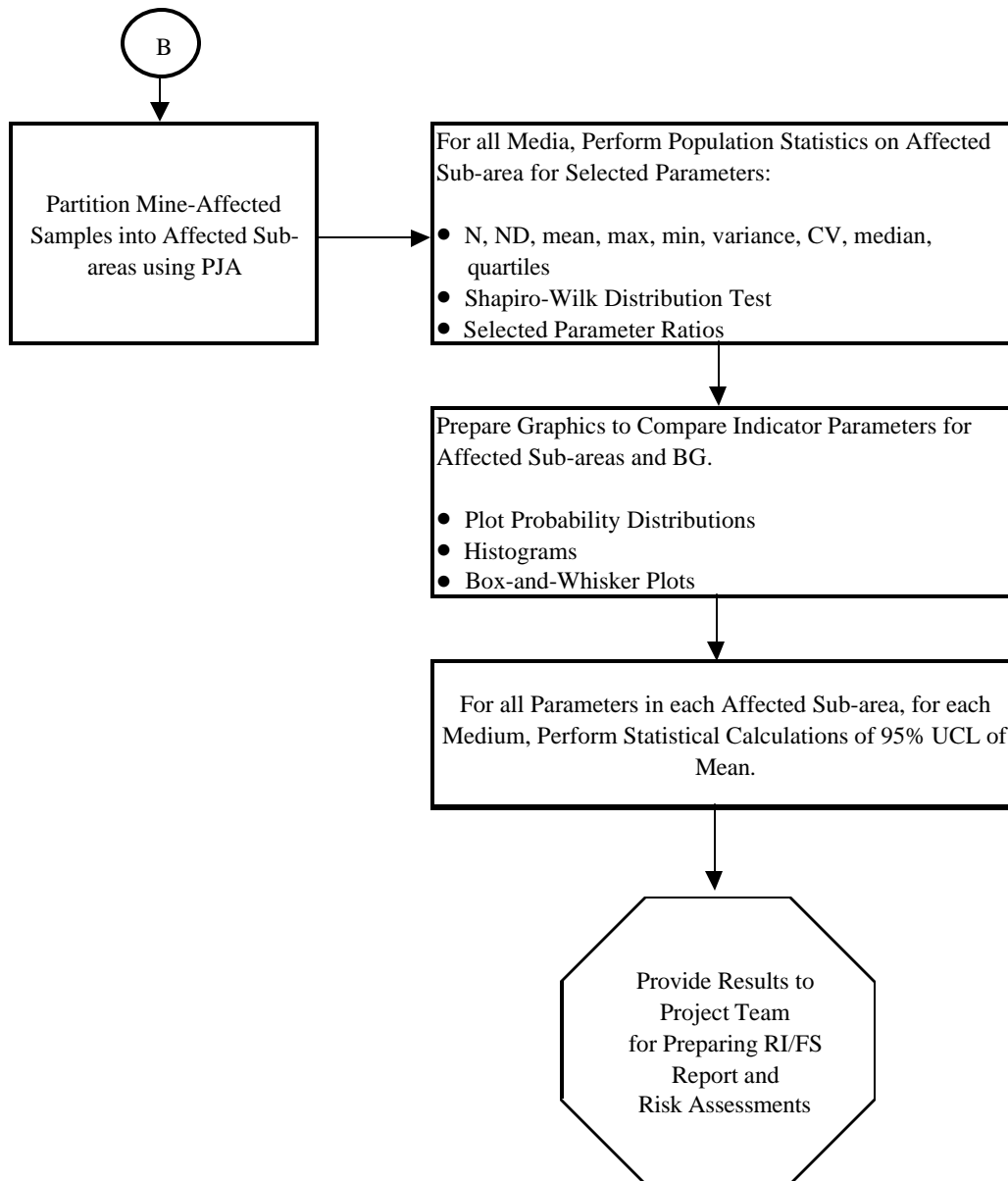


Figure 3: Logic Flow Diagram 3

2.0 EVALUATION OF BACKGROUND DATA

The purpose of the statistical methods described in this section is to accurately estimate background levels of natural chemical constituents in media unaffected by mining activities. Background locations were sampled from natural areas near the site that have analogous geological, geochemical, and hydrogeological conditions to those in the MA and PIA. The MA and PIA include both mineralized and non-mineralized rock types, which was also true prior to mining.

As stated in the Phase 1A QAPP, statistical analyses of concentrations from the BG samples will be used to define the BL values for each medium.

For this study, the term “background limit (BL)” is defined to mean the estimated upper limit of natural background constituent concentrations (Section 2.3). A short list of indicator parameters will be selected for use in comparing site data to the background limits (Section 2.4). Two sets of upper limits will be determined, the background limit (BL) and the re-test background limit (RBL). Next, in the first step comparison, the BL will be compared to PIA sample data. For PIA samples showing exceedances of BL for indicator parameters, the comparison of PIA data to BG data will involve another step. For samples showing exceedances, the second step will involve comparing the re-test BL to data from verification re-samples.

The method used to determine BL and RBL values depends on the population distribution of the data for each parameter, which may be parametric (normal or log-normal) or non-parametric. According to EPA (1992a, 1998), non-parametric analysis is a term referring to a statistical technique applied to analyze the sample data that do not follow a specific distribution (such as normal or log-normal). Parametric is a term referring to a statistical technique to analyze sample data that can be assumed to follow a specific distribution (e.g., normal or log-normal). Verification re-sample means a new, independent sample collected from the same location and analyzed for the same constituent that exceeded the BL. This methodology is further described in the following sections of this document.

The approach planned for this project is consistent with available EPA guidance for statistical methods to detect the presence of contamination (EPA 1989) and other guidance documents on statistical methods (ASTM 1996). Our planned approach allows an evaluation of whether individual locations in the PIA are affected by mining activities rather than simply comparing average values in the BG samples to those in the PIA. Simple comparisons between average values of background and PIA data populations may be misleading because if unaffected PIA samples are pooled with affected PIA samples, this would cause the average value for the PIA population as a whole to decrease relative to the average of only the affected samples. Thus, our approach will be more definitive in identifying the presence and extent of the affected media.

A series of steps will be taken to discriminate areas affected by historical mining. Initially, the BG data will be partitioned by medium (e.g., GW, SW, SED, and SM samples). In some instances, the data for a given medium will be partitioned into more refined populations to reflect the presence of distinct hydrogeological units, such as groundwater in alluvium and bedrock. For sediments,

samples will be grouped by sample type, either as grab (single point) samples or composite (multi-point) samples. In other instances, data for a given medium will be combined into larger groups (e.g., GW and SW data from samples in different seasons). For each medium/data group, basic statistics and statistical tests will be performed (e.g., Shapiro-Wilk Test). From these background data, BLs will be determined as described in Section 2.3.

As illustrated on Figure 1, specific activities for compilation and evaluation of background data will include:

- Select groups of data on which to perform the statistical calculations for each medium (e.g., GW, SW, SED, and SM sample data).
- For each media/data group, use a statistical program to calculate basic statistics and perform tests (e.g., Shapiro-Wilk Test) for normal and log-normal distributions. (See example output in Table 1.)
- Modify statistical program to accommodate negative input values in performing log transformations and statistical calculations. [Note: For this project, the calculation spreadsheet with input/output and formulae with the modification will be made available to interested parties.]
- Use statistical program to calculate the type of data distribution for each group/parameter and univariate statistics, including the 99% upper tolerance limits (UTL) and the 95% upper prediction limits (UPL).
- For each medium, reduce the full suite of analytical parameters to approximately ten indicator parameters based on parameter correlation values for the combined PIA and MA data set and professional judgement.
- Plot histograms and probability-plots for indicator parameters for each data group.
- Determine BL values for each measured parameter and RBL values for each indicator parameter.

The rationale for the statistical analyses and methodology identified on Logic Flow Diagram 1 (Figure 1) are further described in the following sections.

2.1 Grouping of Data by Medium

The following section provides a brief discussion of how the data will be grouped within each medium—highlighting how key issues will be addressed.

2.1.1 Groundwater

Parameters have been measured on samples from GW wells in BG areas unaffected by mining that have analogous mineralized and non-mineralized rock types and similar hydrogeological conditions to those in the MA and the PIA. The MA and PIA sampling locations, which also include

mineralized and non-mineralized rock types, were selected to measure constituent concentrations in the site vicinity under current post-mining conditions. Concentrations of constituents in GW are related to the mineral types and populations in the rock and sediments contacted by the groundwater and the groundwater residence time with those materials. For this reason, the BG alluvial groundwater contains substantially different constituent concentrations than the BG bedrock groundwater. There are a sufficient number of BG samples so that GW data may be partitioned into separate groups for each of the two hydrogeologic units: alluvium and bedrock. Therefore, for this study, all BG GW data will be partitioned into separate alluvium and bedrock populations for the statistical calculations.

Preliminary statistical analysis of BG GW data from Phase 1A indicate that there may be differences between fall and spring sampling rounds for some parameters. However, this seasonal effect does not appear to be significant for many parameters. Currently the number of data points available for each seasonal sampling round may be too small to meet the project goals for statistical confidence and power in comparing PIA to BL values if the data were partitioned into fall and spring groups. Thus, for groundwater it is desirable to pool the data from the different seasonal sampling rounds together into one group. However, it is important to better understand the seasonal effects on the indicator parameters (see Section 2.4 for selection of indicator parameters). Therefore, we will apply graphical analysis methods, such as probability plots and histograms, to evaluate the potential differences between the indicator parameter populations for each season prior to pooling the data. For the indicator parameters, we will apply professional judgement in deciding whether to pool the data for all URS GW sampling rounds. Our goal is to pool the data across seasons for all the subsequent statistical analyses, unless doing so reduces our confidence in discriminating affected PIA sample locations from unaffected locations.

2.1.2 Surface Water

Preliminary statistical analysis of the Phase 1A BG SW data indicated that significant differences exist for some parameters for fall and spring sampling rounds. However, as for the GW data, there may be insufficient seasonal sample events to maintain a reasonable power and confidence if the SW data were to be partitioned by season. Thus, it is desirable to pool together the data from the different seasonal sampling rounds. However, it is important to completely understand the seasonal effects on the indicator parameters prior to pooling the data. Therefore, we will apply graphical analysis methods, such as probability plots and histograms, to evaluate the potential differences between the indicator parameter populations for each season. For the indicator parameters, we will apply professional judgement in deciding whether to pool the data for all URS surface water sampling rounds. Our goal is to pool the SW data across seasons for all the subsequent statistical analyses, unless doing so reduces our confidence in discriminating affected PIA sample locations from unaffected locations.

2.1.3 Sediment

Historically, sediment data have been collected both as discrete point samples and as composite samples. Composite data are not comparable to discrete samples. To appropriately compare recent URS data and historical data, sediment data from point samples will be grouped separately from the composite SED samples data for the statistical analyses.

2.1.4 Surface Material

PIA areas downwind of the MA may have been affected by deposition of windblown dust from the MA. In addition, mechanical transportation of MA material may have occurred along MA haul roads. As described in the Phase 2A/1B QAPP, data from the two BG surface material sub-areas (the mineralized area and the non-mineralized area), will be pooled to reflect the diversity of surface materials overlying geologic units that are analogous to those existing in the MA prior to mining. However, whether the sample data from the two sampled depths should be pooled depends on further analysis of population statistics. Friable surface material samples collected at depths of 0 to 5 cm below the surface litter (shallow) and 5-20 cm (subsurface) have been taken from several PIA sub-areas to evaluate potential effects. An analysis of variance (ANOVA) study will be done to determine if a significant depth effect exists in the BG samples. If there is a significant effect, the SM data will be separated into the two depth categories. If not, the data will be pooled into a single data set for the statistical analyses. The preliminary ANOVA results will be provided to EPA for review and discussion.

2.2 Handling Parameter Values Below Detection Limit

The basis of any parametric statistical interval limit requires the estimation of the sample mean and standard deviation (see Equations 1 and 3). If any value is below its detection limits, then by definition its non-detect concentration/activity is unknown. There is no widely used method that gives good estimates of variance and mean values when many non-detects are present. The most common method of assigning arbitrary values to the non-detects such as one-half the detection limit may result in unreliable estimates of statistical parameters such as the sample mean and variance. One method that sidesteps this last issue is the Helsel's Robust Method described in the next section.

2.2.1 Helsel's Robust Method Applied to Samples Below the Detection Limit

The concentration of a parameter, which is reported by the laboratory as being below the detection limit, is referred to as a non-detect and generally given data qualifiers "U" or "UJ." Careful consideration is given to BG non-detects before calculating tolerance and prediction limits. No statistical analysis is reliable when more than 90% of the samples are non-detects. For those BG data with a detection frequency of greater than 10%, the statistical method described below will be used. Helsel and Hirsch (1992) evaluated several methods for handling non-detect data including substitution, distributional, and robust methods. They concluded that robust methods consistently produced smaller errors when estimating summary statistics, even when multiple reporting limits

were used for non-detect values in the data set. They developed a robust method that combined measured data above the laboratory reporting limit with extrapolated values for non-detect values. The measured and extrapolated values were used collectively to estimate summary statistics.

Helsel's Robust Method will be applied for estimating the value of each non-radiological chemical constituent that is reported to be below detection limit, which we will treat as a non-detect value (Helsel and Hirsch 1992). For each sample analysis showing such a non-detect value, the same estimated quantity for a non-detect value will be used in all statistical calculations using that sample data. On the other hand, the non-detect values for all radioactivity measurements will not be estimated using Helsel's Robust Method because the radioisotope measurement and reporting methods, which are different from chemical analysis methods, and such non-detects are more appropriately estimated using the instrument readout values. Therefore, any radioisotope value flagged in the project database as a non-detect that is greater than zero will be used without adjustment. However, if a radioisotope value is reported by the laboratory as a zero or a negative value, then the existing statistical programs must be modified. The purpose of this modification is to appropriately accommodate radioactivity values that are below detection limits but are reported by the laboratory as the measurement result rather than as a non-detect with an associated detection limit. In the current project Technical Data Management System (TDMS) data set, radiological data were reported with two separate parameter values: "2-sigma" and "ES." For the statistical calculations on radiological data, values below detection limits will be treated differently than for the chemical laboratory analyses, as described below.

2.2.2 2-Sigma Values from Radiological Laboratory Analyses

The 2-sigma values (listed under the column labeled "2Q" in the project database) represent the 2-sigma counting error value associated with the laboratory analysis. This type of radiological data is one measure of the uncertainty in the measurement and is used during validation as one means of identifying non-detects, which are then flagged with a U or UJ qualifier. However, neither these 2-sigma values nor the detection limit values will be used in calculating statistical background limits. Instead, the statistical calculations will use the laboratory measurement values for each radiological analysis, which are listed in the column ES in the project database.

2.2.3 Negative ES Values from Radiological Laboratory Analyses

The ES data may contain a zero or a negative value. These data are considered non-detectable values and are flagged by U and UJ qualifiers. For arithmetic statistics and testing for normal distributions, the reported values, including the negative results will be used. Any resultant negative statistic, such as a mean, mode, UTL or UPL will be set to zero. For statistics on log-transformed data, a small constant will be added to all data to eliminate any zero or negative value. This additive constant is generally called a third parameter in log-normal theory (Crow and Shimizu [Ed] 1988). The log-normal distributional test can be done on the adjusted values and give meaningful results. However, before reporting any statistics such as a mean or limit, the additive constant will be removed.

2.3 Univariate Statistics and Distribution Testing

Basic univariate statistics and a distribution test will be done on the background data.

2.3.1 Basic Statistics

As stated previously, the term “background limit” is defined to mean the estimated upper limit of natural background constituent concentrations. The method used to estimate the BL depends on the population distribution of the available BG data for the parameter (e.g., normal, log-normal, non-parametric distributions), and the frequency of laboratory detected values for that parameter. If there is an insufficient number of detected values in a data set for a given parameter (i.e., less than 10%) to calculate a meaningful statistically-based limit, the BL will not be determined statistically. Rather, the BL will be set equal to the maximum measured value, if there are sufficient detected values. In the unlikely case in which the laboratory data for a given parameter show no detected values, the BL will be set equal to the laboratory quantitation limit (QL).

As shown on Figure 1, if the data for a given parameter fit a normal or log-normal distribution, then the BL will be set equal to the parametric UTL. In simple terms, the UTL means the statistically-calculated estimate of the upper end of BG data for a parameter with a given expected coverage (say 99%) such that 99% of non-impacted site sample values are expected to be less than that UTL value. If the data for a given parameter do not fit either a normal or log-normal distribution, then a non-parametric UTL will be calculated. In simple terms, the 95% UPL means a statistically-calculated prediction of the upper end value of the BG data for which the next single site sample value has a 95% chance of falling below the UPL value if the sample is not mine affected. For the upcoming data evaluation, we have selected the 99% expected coverage for each UTL and the 95% confidence interval for each UPL, consistent with the sampling and statistical comparison strategy recommended in EPA guidance (EPA 1992a). These confidence intervals are recommended as a means to balancing false positive and false-negative errors in making the comparisons between BG and PIA data. As stated in the Phase 1 QAPP for this RI/FS project (URS Greiner 1999), one overall statistical goal for this study is a site-wide false-positive error (α) of 0.05. Another goal is to achieve a low false-negative error (β) that is consistent with EPA guidance (EPA 1992a).

Anticipating the large number of comparisons of BL values to site data, we have incorporated verification sampling and a two-step statistical testing strategy, which is consistent with a recommended strategy for multiple comparisons in the EPA guidance for statistical evaluation of groundwater monitoring data (EPA 2000). Compared to other documents we are aware of, this EPA groundwater guidance provides the most sophisticated description for strategies for balancing false-positive and false-negative errors. The importance of balancing false-positive errors is also noted in the EPA guidance on sampling within soil and solid media (EPA 1994). Also, the EPA guidance for data quality assessment emphasizes the mathematical principles for achieving a balance in these potential errors with reference to any particular medium. Thus, we have adopted a strategy

recommended in the EPA groundwater guidance (EPA 2000) for multiple comparisons of samples of the other media as well as for groundwater samples.

Implementing this strategy involved sampling of numerous background locations, determining BL values, and collecting verification samples for each medium at each PIA location showing values exceeding a preliminary BL for potential indicator parameters. Subsequently, after the complete validated data set is available, we will perform a two-step comparison test of indicator parameters. The rationale for selecting indicator parameters is described further in the following sections.

In the first-step test we will compare PIA data to the BL (UTL) for indicator parameters. For PIA samples exceeding BL values, a second-step test will be performed using an RBL. As shown in the lower portion of Figure 1, the RBL for each indicator parameter will be determined using a similar rationale as described above for determining the BL, except that the RBL values will be based on the upper prediction limit at a 95% confidence interval (i.e., 95% UPL). The use of a higher UTL level and a somewhat lower UPL level is consistent with recommendations in the EPA guidance (1992a). A more rigorous statistical definition of the UTL and UPL is provided below in Sections 2.3.1.1 and 2.3.1.2.

Table 1 presents an example output of some univariate statistics that will be determined and tabulated. Columns from left to right are:

1. Parameter name (**Parameter**).
2. Number of samples (**N**).
3. Number of non-detects (**ND**).
4. Minimum (**Min**) value measured in sample data set. (The lowest estimated concentration used to calculate the 99% UTL is provided in parentheses. For non-radiological parameters, Helsel's distributional method is used to replace non-detects.)
5. Maximum (**Max**) value measured in sample data set.
6. Probability that the sample data represents a normal distribution (**Normal Prob**). A probability of 0.05 or less indicates that it is unlikely that the data is from a normal distribution.
7. Probability that the sample data represents a log-normal distribution (**Log-normal Prob**). A probability of 0.05 or less indicates that it is unlikely that the data is from a log-normal distribution.
8. **Method**. If the number of non-detects is high (i.e., 90% or more) a statistical approach is not applicable and the maximum concentration, or laboratory quantitation limit if all samples are non-detect, is used as the limit. If the probabilities for both the normal and log-normal distribution are 0.05 or less, a non-parametric method is used to estimate the 99% UTL (95% UPL). Otherwise, the distribution with the highest probability is used to calculate the 99% UTL (and the 95% UPL), which are then the BL and the RBL values.

9. Background limit (BL). On the example Table 1, the BL is labeled **preliminary background limit**. Values have been rounded to the appropriate number of significant figures.
10. **Re-test background limit (RBL)**. If there is no exceedance of BL for a given parameter, then the RBL is not used for comparison to verification samples. Values have been rounded to the appropriate number of significant figures.
11. Units of measure (**Units**).

After these statistical analyses are complete, we plan to provide EPA a working draft of the tabulated results for their review.

Midnite Mine Surface Water 95% Tolerance and 95% Prediction Limits Analysis										
Parameter	N	ND	Min	Max	Normal Prob	Log Normal Prob	Method	Preliminary Background Limit	Retest Background Limit	Units
Alkalinity, Total	48	0	11.4	214	0.000	0.028	nonparametric	214	214	ug/l
Aluminum, dissolved	48	1	3.5483	8620	0.000	0.062	lognormal	9501.5	9501.5	ug/l
Aluminum, total	48	2	21.1	9540	0.000	0.001	nonparametric	9540	9540	ug/l
Antimony, dissolved	48	35	0.9006	6.2	0.000	0.028	nonparametric	6.2	6.2	ug/l
Antimony, total	48	31	0.5292	5.5	0.000	0.595	lognormal	4.7296	4.7296	ug/l
Arsenic, dissolved	48	0	0.51	21.3	0.000	0.938	lognormal	12.263	12.263	ug/l
Arsenic, total	48	0	0.57	46.1	0.000	0.816	lognormal	20.062	20.062	ug/l
Barium, dissolved	48	0	1.9	252	0.000	0.013	nonparametric	252	252	ug/l
Barium, total	48	0	2.7	274	0.000	0.148	lognormal	138.01	138.01	ug/l
Beryllium, dissolved	48	39	0.1	0.26	0.000	0.000	nonstatistical	0.26	0.26	ug/l
Beryllium, total	48	26	0.0242	0.5	0.000	0.284	lognormal	0.3293	0.3293	ug/l
Bicarbonate Alkalinity	48	0	11.4	214	0.000	0.028	nonparametric	214	214	ug/l
Cadmium, dissolved	48	40	0.2	0.5	0.000	0.000	nonstatistical	0.5	0.5	ug/l
Cadmium, total	48	39	0.2	0.25	0.000	0.000	nonstatistical	0.25	0.25	ug/l
Calcium, dissolved	48	0	1550	63300	0.000	0.189	lognormal	29915	29915	ug/l
Calcium, total	48	0	2080	62800	0.000	0.079	lognormal	29694	29694	ug/l
Carbonate Alkalinity	48	40	2	2	0.000	0.000	nonstatistical	2	2	ug/l
Chemical Oxygen Demand	48	21	0.6225	41.9	0.000	0.721	lognormal	33.844	33.844	ug/l
Chloride	48	0	0.525	2.49	0.000	0.001	nonparametric	2.49	2.49	ug/l
Chromium, dissolved	48	31	0.0408	3.2	0.000	0.456	lognormal	2.07	2.07	ug/l
Chromium, total	48	19	0.2518	7.2	0.000	0.009	nonparametric	7.2	7.2	ug/l

Table 1: Example .o1 file from SAS output

2.3.1.1 Upper Tolerance Limit

The tolerance limit method suggested in the EPA guidance document (EPA 1989), will be used to establish the first-stage BL of each measured parameter. A tolerance limit is the upper bound of a statistical interval calculated to include, on the average, a specified proportion of future observations from the same population. This proportion is also referred to as the average coverage of the tolerance limit. An average coverage of over 90% is the target for this study.

The UTL is defined as:

$$UTL = m + 1s$$

Equation 1: UTL

Where:

- μ = mean value of parameters in data set
- σ = standard deviation of parameters in data set

$$I = t_{(n-1, 1-\alpha)} \sqrt{1 + 1/n}$$

Equation 2: I for UTL

- n = number of background samples
- t = t-distribution
- α = significance level or false-positive rate for the individual comparison (that is, one sample and constituent or parameter)

For non-parametric parameters, the UTL is generally set to the maximum value found in the BG data (Davis and McNichols 1994a).

2.3.1.2 Upper Prediction Limit

A prediction limit is the upper bound of a statistical interval calculated to include verification samples from the same population with a specified confidence level, i.e., $1-\alpha$. When more than one parameter is being rechecked, a Bonferroni correction will be employed (down to a minimum α of 0.01) (EPA 1992a). For the k indicator parameters that exceed UTLs, we plan to collect a verification sample and re-analyze the parameter. If any of those analyses exceeds the UPL, then the initial UTL exceedance is considered confirmed, and the sample is provisionally categorized as “affected” by mining. Otherwise, all of the initial exceedances are considered to be false-positives and the initial sample is categorized as “non-affected” by mining. If the background is parametric, then the definition of an UPL is:

$$UPL = \bar{m} + I s$$

Equation 3: UPL

Where:

- μ = mean value of parameters in data set
- σ = standard deviation of parameters in data set

$$I = t_{(n-1, 1-\alpha/k)} \sqrt{1 + 1/n}$$

Equation 4: I for UPL

- k = number of parameters re-sampled for verification
- n = number of background samples
- t = t-distribution
- α = significance level or false positive rate for the individual comparison (that is, one sample and constituent or parameter)

Table 2 shows the multiplication factor λ for a 99% UTL with various sizes of background data sets. This table shows the “experimentwise” 95% UPL for various k UTL exceedances in a sample. Experimentwise means that the total test (experiment) is adjusted so that the overall false-positive error rate is maintained. The value k is defined as the number of parameters re-sampled for verification. Pursuant to EPA guidance, the column with k = 5 should be used if the number of UTL exceedances are greater than or equal to 5. In a case where a verification re-sample is collected and analyzed for other parameters than those exceeding the BL, the verification re-sample data for the other parameters will not be used for comparison to a BL or RBL.

If the distribution is non-parametric, then the UPL is defined as the maximum value (Davis and McNichols 1994b; EPA 1992a).

Table 2: Chart with factor lambda (λ) for both a BL (99% UTL) and a RBL (95% UPL)

Background Size	99% UTL	----- lambda at 95% UPL-----					
n		k=	1	2	3	4	5
4	5.08		2.63	3.56	4.18	4.67	5.08
5	4.10		2.34	3.04	3.49	3.83	4.10
6	3.63		2.18	2.78	3.15	3.42	3.63
7	3.36		2.08	2.62	2.94	3.17	3.36
8	3.18		2.01	2.51	2.80	3.01	3.18
9	3.05		1.96	2.43	2.70	2.90	3.05
10	2.96		1.92	2.37	2.63	2.82	2.96
11	2.89		1.89	2.33	2.58	2.75	2.89
12	2.83		1.87	2.29	2.53	2.70	2.83
13	2.78		1.85	2.26	2.49	2.66	2.78
14	2.74		1.83	2.24	2.46	2.62	2.74
15	2.71		1.82	2.22	2.44	2.59	2.71
16	2.68		1.81	2.20	2.42	2.57	2.68
17	2.66		1.80	2.18	2.40	2.54	2.66
18	2.64		1.79	2.17	2.38	2.53	2.64
19	2.62		1.78	2.16	2.36	2.51	2.62
20	2.60		1.77	2.14	2.35	2.49	2.60
21	2.59		1.77	2.14	2.34	2.48	2.59
22	2.57		1.76	2.13	2.33	2.47	2.57
23	2.56		1.75	2.12	2.32	2.46	2.56
24	2.55		1.75	2.11	2.31	2.45	2.55
25	2.54		1.74	2.10	2.30	2.44	2.54
26	2.53		1.74	2.10	2.30	2.43	2.53
27	2.52		1.74	2.09	2.29	2.42	2.52
28	2.52		1.73	2.09	2.28	2.42	2.52
29	2.51		1.73	2.08	2.28	2.41	2.51
30	2.50		1.73	2.08	2.27	2.40	2.50
31	2.50		1.72	2.07	2.27	2.40	2.50
32	2.49		1.72	2.07	2.26	2.39	2.49
33	2.49		1.72	2.07	2.26	2.39	2.49
34	2.48		1.72	2.06	2.25	2.38	2.48
35	2.48		1.71	2.06	2.25	2.38	2.48
36	2.47		1.71	2.06	2.25	2.37	2.47
37	2.47		1.71	2.06	2.24	2.37	2.47
38	2.46		1.71	2.05	2.24	2.37	2.46
39	2.46		1.71	2.05	2.24	2.36	2.46
40	2.46		1.71	2.05	2.23	2.36	2.46

Note: (k= number of previous UTL exceedances) given the number of background samples (n). The values in Table 2 are based on one comparison to the RBL.

2.3.2 Shapiro-Wilk Distribution Test

Whether or not the population distribution of a data set is normal, log-normal, or non-parametric will be established using the Shapiro-Wilk's W test. The Shapiro-Wilk distribution test is one of the most powerful tests for normality (or log-normality). This test is similar to computing a correlation between the quantiles of the standard normal distribution and the ordered values of the data. The Shapiro-Wilk W statistic tends to be close to 1.0 when a probability plot of the data indicates a nearly straight line.

Tolerance and prediction limits are calculated differently if the data have normal or log-normal parametric or non-parametric distributions. The calculations of the tolerance and prediction limits are carried out for each parameter in the scale (e.g., arithmetic or logarithmic) that has the greater probability of being normally distributed. If the probability of normality is less than 5%, a non-parametric method will be used for determining both background limits (EPA 1989). Both these limits are the maximum value of all the sample concentrations for the parameter.

2.4 Selection of Indicator Parameters

Over one hundred different parameters have been measured on samples collected for this RI/FS project, including major ions, metals (total and dissolved), and GW parameters such as pH and multiple radioisotopes for some elements (such as uranium 235). Comparing many samples in the PIA to BL values for all the measured parameters would create an unreasonably high false-positive error rate for the site (URS Greiner 1999). Thus, to maintain a reasonable false-positive error rate in balance with the false-negative error rate, the parameter list for the BG/PIA data comparisons must be reduced to a more manageable number (EPA 1992a, Gibbons 1991, McNichols and Davis 1998). Therefore, a reduced set of the measured parameters, which are referred to herein as indicator parameters, will be identified for each medium and used for the statistical analyses to discriminate and identify mine-affected samples. As previously described in the Phase 1A QAPP, approximately 10 indicator parameters will be identified, which are potentially indicators of mine-related contamination, for each medium.

We plan to provide a list of recommended indicator parameters to EPA for review and approval prior to performing the comparisons of BL values to PIA sample data. (The methodology is discussed in greater detail in Section 4.0.)

2.4.1 Correlation Coefficients and Matrix

It is desirable to select indicator parameters that are not correlated with one another because this will reduce the false-positive error. To assist in selecting the appropriate indicator parameters for each medium, a correlation matrix of all parameters measured in the PIA and MA samples will be determined and printed in a matrix form to illustrate the extent of correlation between prospective indicator parameters. A working draft of this correlation matrix will be provided to EPA for review and discussion during the process of selecting indicator parameters.

2.4.2 Professional Judgement Assessment

In addition to evaluating the correlation matrix, professional judgement and experience in mine-related contamination will be applied to appropriately select the indicator parameters. The selection of indicator parameters will consider geochemical and hydrogeological processes and previous experience in mining-related contaminants and migration.

3.0 COMPARISON OF PIA SAMPLES TO BACKGROUND LEVELS

The purpose of comparing the value of each indicator parameter from an individual PIA sample to a BL is to evaluate whether individual locations in the PIA are affected by mining activities. This section describes the methods for comparing PIA sample data to statistically-based background limits and other statistical analyses that will be used to augment a professional judgement in making final determinations of which PIA samples have been affected by the mining activities (Koch and Link 1970).

The method used to compare individual PIA samples to BL values will be based on whether a parametric or non-parametric statistical method will be used. If a parametric comparison is conducted, then a UTL or UPL will be calculated using Equations 1 and Equation 3, respectively. Non-parametric comparisons require that either a maximum value or a laboratory quantitation limit be used for that parameter.

As illustrated in Figure 2, specific activities for comparison of PIA samples to BL values will include:

- For each medium, compare indicator parameter values for each PIA sample to the appropriate BL.
- Determine if PIA indicator parameters exceed BL values.
- If no PIA samples show indicator parameters exceeding the BL test, assign that sample to non-affected class.
- For PIA samples having indicator parameters that exceed the BL test, determine if PIA verification sample exceeds RBL for each indicator parameter that exceeded its BL.
- If all the previous PIA exceedance parameters do not exceed their RBL value, assign that sample to non-affected class.
- If any verification sample shows an exceedance of the RBL, assign sample to the “provisionally” affected class.
- Perform multivariate graphical assessment for all media in the provisionally affected class data
- Perform the Mann-Whitney “U” test for surface material indicator parameters to determine if the provisionally affected class data is significantly different from the BG data.
- Using results of the above activities and professional judgement, make final assignment of samples to affected and non-affected classes and plot exceedance wheels at sample locations.

3.1 Exceedance Wheel Plots

Results of this two-step test can be illustrated using a wheel-like graphic, referred to as an “exceedance wheel” (EW) plot. These EW plots will also provide graphic support for the elements of the RI/FS that follow the activities described herein. On the EW plot, the radius of the “rim” of

the wheel is standardized to unity to represent each indicator parameter's BL or RBL. Each spoke of the wheel represents an individual PIA sample indicator parameter value whose length is made relative to its respective BL or RBL. Once the final validated data set is available for each indicator parameter for each medium, a final BG (e.g., 99% UTLs) will be calculated. As described earlier in this section, the first step of this test will be a comparison of the PIA sample data to the 99% UTL. In the example shown in Figure 4, parameters 3 and 7 have exceeded their respective 99% UTLs¹. The sampling strategy developed for this project (URS Greiner 1999) included collecting verification re-samples at sample locations showing initial exceedances of preliminary BLs (UTLs).

For the second step of the test, the data from the verification samples will be compared to the RBL (UPL) and illustrated by the EW plot. For samples showing exceedances of preliminary BLs, verification sampling was conducted. Figure 4 is the UPL EW using Table 2 again. Given two initial exceedances, two re-tests will be required ($k=2$). Given that the number of BG samples is 32, the UPL λ value of 2.07 can be found again on the row labeled 31 and in the column labeled $k=2$. Equation 3 can now be used to calculate the UPL for each indicator parameter tested. The EW plot in Figure 4 shows these results as two twinned spokes representing both the initial BL exceedances and the verification sampling data compared to the RBL values. In this case, with no UPL exceedances, the sample location can be assigned to the non-affected class.

To facilitate the professional judgement assessment in discriminating mine-affected samples, an EW plot will be created for each sample. For samples having an exceedance of a background limit for any indicator parameter, the EWs will be plotted on site maps for each medium to allow visualization of areal trends of elevated values. For samples that do not show exceedances of BL values, the EWs will be created and made available for viewing separately by interested parties, but not plotted on the site maps to avoid creating excessive graphical clutter.

¹ For this example, assume that the background data set had 32 samples and all indicator parameters have normal distributions. This produces a UTL λ equal to 2.5. This can be read from Table 2 at the row labeled 31 ($n-1$) and the column labeled 99% UTL. Equation 1 can now be used to calculate the UTL for each of the indicator parameters.

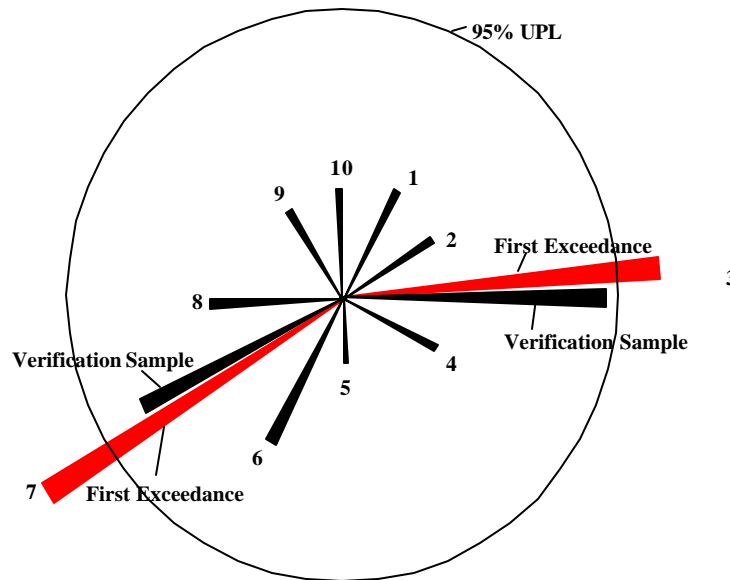


Figure 4: Example Exceedance Wheel Plot Showing 95% UPL for water samples with $k = 2$ UTL exceedances, but no UPL exceedance.

1. Aluminum
2. Antimony
3. Chromium
4. Cobalt
5. Iron
6. Manganese
7. Nickel
8. Sulfate
9. Uranium-234
10. Zinc

Figure 5 presents another possibility. Again, parameters 3 and 7 have exceeded their respective 99% UTL and again there are 32 BG samples. The same values for UTL and UPL are calculated. In this case, Parameter 3 has exceeded its 95% UPL. This implies that the sample can provisionally be assigned to an affected population.

It should be mentioned again that if any indicator parameter is log-normal, all calculations are done with log transformed values. If any indicator parameter is non-parametric, both sample and possible verification sample results will be tested against the maximum BG value for that parameter.

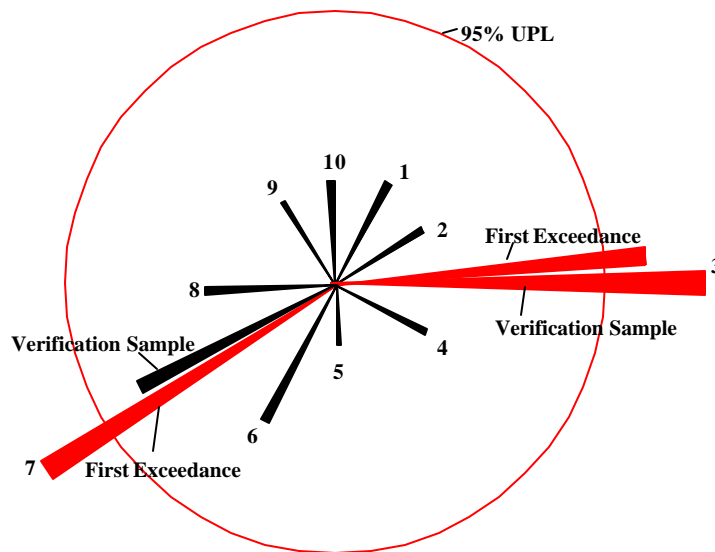


Figure 5: Example Exceedance Wheel Plot Showing 95% UPL for water samples which have $k = 2$ UTL exceedances and 1 UPL exceedance.

1. Aluminum
2. Antimony
3. Chromium
4. Cobalt
5. Iron
6. Manganese
7. Nickel
8. Sulfate
9. Uranium-234
10. Zinc

3.2 Mann-Whitney “U” for Surface Material Indicator Parameters

The Mann-Whitney U or Wilcoxon Rank Sum test is a non-parametric analog to the population-to-population t-test. As described in the Midnite Mine Phase 2A/1B RI/FS QAPP, this Phase 2 statistical approach will be used for surface material samples only, to evaluate whether or not the average concentrations in affected sub-areas exceed BG. This is an approach recommended in MARSSIM (2000). Only the samples within the affected sub-areas, which will be defined areally by professional judgement in evaluating the data, will be included in the PIA data set for comparing each affected sub-area population to the background data population. A working draft of the results of this analysis will be provided to EPA for review and discussion.

3.3 Multivariate Graphical Assessment

Graphical displays showing various views of the multivariate nature of BG and PIA sample data may help classify a sample as affected or unaffected by mining. The graphics may include multiple box-and-whisker plots, ternary, 3-D scatter plots, etc. At this stage, parameters other than the 10 indicators may be included. A working draft of the results of this analysis will be provided to EPA for review and discussion.

3.4 Final Discrimination and Plotting Locations

A final decision will be made on the affected and non-affected classification of individual samples. This step, which is the final step in discriminating mine-affected samples from background, will utilize professional judgement assessment (PJA) in integrating and possibly reconciling results from the previous steps. The PJA will consider such factors as:

- Number and levels of exceedances
- Areal trends relative to source locations and migration processes
- Characteristics of contamination at other mine sites with acid rock drainage, metals leaching, and migration processes

Prior to completing the draft technical memo on the results of the full set of statistical analyses, the results of this evaluation will be provided to EPA—as a working draft for review and discussion purposes.

4.0 STATISTICAL CHARACTERIZATION OF MINE-AFFECTED SUB-AREA POPULATIONS

This section describes the methods that will be used to organize mine-affected samples into populations and delineate the areas containing samples showing effects of mining activities. The purpose of this evaluation is to characterize the constituent levels in the mine-affected sub-areas of the PIA for all analyzed parameters and provide the statistical results to the project team for use in:

- Characterizing the nature and extent of contamination
- Preparing the RI report
- Performing the risk assessments
- Preparing FS estimates of volumes of material needing remediation

The specific steps of this evaluation are:

- Create groups of mine-affected samples and delineate each affected PIA sub-area for each medium. Sample grouping will consider the same factors noted above for the PJA.
- Determine population statistics for each affected sub-area for each medium. The statistics include max, min, n, number of non-detects, etc., and Shapiro-Wilk distribution tests.
- Prepare graphics (e.g., histograms and box-and-whisker plots) to compare the indicator parameters for the mine-affected sub-area to BG values.
- Determine the 95% UCL of the mean for each mine-affected sub-area using Equation 5 and EPA-recommended methods (EPA 1997).
- Provide tables containing 95% UCL values and population statistics for each parameter measured in each medium, in each sub-area to the project team for use in preparing the RI/FS.

4.1 Calculation of Univariate Statistics

As shown in the top-middle box on Figure 3, basic population statistics will be determined for the mine-affected sub-areas. These results will provide input for the RI report and the risk assessments. During the risk assessment process, additional statistical analyses will be performed for that purpose (e.g., for exposure concentrations in affected areas).

4.2 Graphical Comparisons of Populations

As shown in the second box from the top in the middle of Figure 3, graphical plots will be prepared to illustrate the data populations for the mine-affected sub-areas. The statistical plots will help illustrate and support conclusions derived from the sub-area population. EPA recommends several graphical techniques, including using probability distributions (EPA 2000, 2001).

4.3 Analysis of Parameter Values in Affected Sub-areas

For each mine-affected sub-area, the 95% UCL for data distributed normally will be determined using Equation 5 (EPA 1992b). For log-normal and non-parametric distributions, jackknife or bootstrap calculations are recommended (EPA 1997).

$$UCL = \bar{m} + t_{(n-1, 1-\alpha)} (s / \sqrt{n}) \quad \text{Equation 5: UCL}$$

n = number of affected sub-area samples

σ = standard deviation of parameter

μ = mean value of data for parameter

t = t-distribution

α = significance level of 0.05

5.0 REFERENCES

ASTM. 1996. Provisional Standard Guide for Developing Appropriate Statistical Approaches for Ground-Water Detection Monitoring Programs, PS 64 - 96, American Society for Testing and Materials.

Crow and Shimizu. (Ed) 1988. Log-normal Distributions, Theory and Applications, Dekker Vol. 88.

Davis, C.B., and R.J. McNichols. 1994a. Ground Water Monitoring Statistics Update: Part I: Progress since 1988. *Ground Water Monitoring and Remediation* 14, No.4.

Davis, C.B., and R.J. McNichols. 1994b. Ground Water Monitoring Statistics Update: Part II: Non-parametric Prediction Limits. *Ground Water Monitoring and Remediation* 14, No.4.

Gibbons. Robert. 1991. Statistical Tolerance Limits for Groundwater Monitoring, Groundwater, Vol. 29, No. 4 – July-August 1991, p. 569.

Helsel, D. R. and Hirsch, R. M. 1992. Statistical Methods in Water Resources. Elsevier Sciences Publishers, Amsterdam.

Koch, G. and Link, R. 1970. Statistical Analysis of Geological Data, Dover Publications, Vol. 1 and 2.

MARSSIM. 2000. Multi-Agency Radiation Survey and Site Investigations Manual, NUREG-1575, Rev. 1, EPA 402-R-97-016, Rev. 1, DOE/EH-0624, Rev. 1.

McNichols and Davis. 1998. Statistical Issues and Problems in Groundwater Detection Monitoring at Hazardous Waste Facilities. *GWMR*, Fall 1998, p. 136 and 137.

U.S. Environmental Protection Agency (EPA). 1989. Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities: Interim Final Guidance. Washington, D.C.: Office of Solid Waste, EPA530-SW-89-026.

EPA. 1992a. Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities: Addendum to Interim Final Guidance. Washington, D.C.: Office of Solid Waste, EPA530-SW-89-026 (Addendum).

EPA. 1992b. Supplemental Guidance to RAGS: Calculating the Concentration Term, Pub. 9285.7-081.

EPA. 1994. Statistical Methods for Evaluating the Attainment of Clean-up Standards for Soil and Solid Media. EPA/230/R-94/004. June.

EPA. 1997. The Log-normal Distribution in Environmental Applications, USEPA Technology Support Center Issue, Singh, Singh, and Engelhardt. EPA/600/R-97/006.

EPA. 2000. Guidance for Data Quality Assessment. EPA QA/G-9, EPA/600/R-96/084. July.

EPA 2001. Guidance for Characterizing Background Chemicals in Soil at Superfund Sites: External Review Draft. Washington, D.C.: Office of Emergency and Remedial Response. EPA 540-R-01-003.

URS Greiner. 1999. Quality Assurance Project Plan for the Midnite Mine Phase 1A RI/FS, Stevens County, Washington, 68-W-98-228, 1800.5.71.003.

URS. 2000. Quality Assurance Project Plan for the Midnite Mine Phase 2A-1B RI/FS, Stevens County, Washington, 18068-W-98-228, 1800.64.